

# Des outils pratiques pour gérer et valoriser vos données de recherche

Formation URFIST Bordeaux  
17 mai 2018

**Atelier 3** : « Décrire et documenter ses données avec des métadonnées standardisées »

Martine Barale  
CIRAD

Chargée d'appui à la gestion et à la valorisation  
des données de la recherche

# Le CIRAD en quelques mots...

Recherche  
agronomique

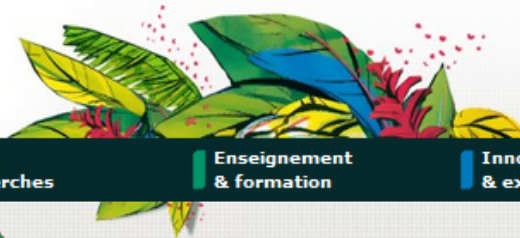
Coopération  
internationale

Développement  
durable

Tropiques et  
Méditerranée

1650 agents,  
dont 800  
chercheurs

Partenariats  
avec plus de  
100 pays sur  
3 continents



- Actualités
- Emplois / Stages
- Contacts / Accès
- English version
- Intranet
- Recherche

Qui  
sommes-nous ?

Nos  
recherches

Enseignement  
& formation

Innovation  
& expertise

Publications  
& ressources



## Le Cirad

L'organisme français de recherche  
agronomique et de coopération  
internationale pour le développement  
durable des régions tropicales et  
méditerranéennes.



Innovons ensemble pour les agricul

[www.cirad.fr](http://www.cirad.fr)

## A la Une



### Une charte de déontologie explicite l'engagement éthique du Cirad

23/04/2018 - Institutionnel

L'éthique constitue un repère fort au sein du Cirad. Après avoir signé en 2015 la charte nationale de la déontologie des métiers de la recherche, le Cirad déploie désormais une charte qui lui est spécifique. Cette charte propose notamment une déclinaison particulière adaptée au partenariat en ...

[Lire la suite](#)



### Pl@ntNet s'ouvre à l'Afrique tropicale

25/04/2018 - Science

1300 espèces de la flore africaine sont désormais identifiables grâce à l'application Pl@ntNet ...

[Lire la suite](#)



### ICOPE 2018 : des solutions scientifiques pour la production d'huile de palme durable

19/04/2018 - Événement

Pour la 6e fois depuis 2007, le Cirad organise avec la société indonésienne Sinarmas et WWF ...

[Lire la suite](#)



### Sara Mercandalli : « Les migrations ont un rôle clé dans la reconfiguration des territoires africains »

12/04/2018 - Questions à



### Dimensions éthiques des grands accords internationaux Objectifs du développement durable et Climat

16/04/2018 - Ça vient de sortir

## Agro-écologie



## Territoires vivants



## Développement durable et filières tropicales

# Le Dataverse du Cirad

 **Ouverture : 30 janvier 2018** Guide Dataverse | **Gérer les Données de la recherche** | User Guide | Support | Log In



CIRAD Dataverse ([dataverse.cirad.fr](https://dataverse.cirad.fr))

Données de recherche du CIRAD

<https://dataverse.cirad.fr/>

Metrics

541 Downloads

Contact Share

Les activités du CIRAD relèvent des sciences du vivant, des sciences sociales et des sciences de l'ingénieur appliquées à l'agriculture, à l'alimentation, à l'environnement et à la gestion des territoires. *CIRAD's activities concern the life sciences, social sciences and engineering sciences, applied to agriculture, the environment and territorial management.*

Search this dataverse...

Find

Advanced Search

☒ Datasets (8)

☒ Files (31)

☐ Files (162)

**Dataverse Category**

Laboratory (7)

Research Group (1)

**Metadata Source**

CIRAD Dataverse (38)

CIRAD (1)

**Publication Date**

2017 (27)

2018 (9)

2016 (3)

**Subject**

Agricultural Sciences (21)

Computer and Information Science (19)

1 to 10 of 39 Results

Sort

Data and code for the publication entitled "The limited contribution of large trees to annual biomass production in an old-growth tropical forest"



Mar 26, 2018 - Forêts et Sociétés

Ligot, Gauthier; Gourlet-Fleury, Sylvie; Ouédraogo, Dakis-Yaoba; Morin, Xavier; Bauwens, Sébastien; Baya, Fidèle; Brostaux, Yves; Doucet, Jean-Louis; Fayolle, Adeline, 2018, "Data and code for the publication entitled "The limited contribution of large trees to annual biomass production in an old-growth tropical forest", doi:10.18167/DVN1/8T4XTV, CIRAD Dataverse, V2

This archive contains all the necessary data and R script to reproduce the results of the manuscript entitled "The limited contribution of large trees to annual biomass production in an old-growth tropical forest" published in Ecological Application journal. It includes three csv...

Arthropods inventory from PELEBINA (BENIN)



Mar 23, 2018 - Aida

Silvie, Pierre; Drieu, Robin; Boucher, Alexandre; Jean, Janine; Auzoux, Sandrine, 2018, "Arthropods inventory from PELEBINA (BENIN)", doi:10.18167/DVN1/WGNKO1, CIRAD Dataverse, V1

The file contains taxonomic data of Arthropods collected during the rainy season (2013) and dry season (first semester 2014) in the rice lowland of Pelebina in Benin (Africa).

Output data from: Combining global tree cover loss data with historical national forest-cover maps to look at six decades of

# CoopIST : des ressources sous licence CC



**CoopIST** Coopérer en information scientifique et technique

Recherche

- Faq
- Contact
- Tous les documents
- Tous les liens

Ressources documentaires

Aide à la publication

Gestion de l'information

A l'international



Gestion de l'information

Gérer les données de la recherche

Gestion de l'information

Agir pour le libre accès

Créer un service

Gérer une bibliothèque

Réaliser des produits

Gérer des identifiants

Gérer les données de la recherche

## Gérer les données de la recherche

Des fiches didactiques sur les concepts et les données sont disponibles sur ce site dans les listes ci-après :

### Gérer les données de la recherche

Découvrir des plans de gestion de données de la recherche

S'initier en ligne aux données de la recherche

Rendre publics ses jeux de données

Découvrir de nouveaux métiers liés aux données de la recherche

Mesurer l'impact des données

Agir pour le libre accès

Créer un service

Gérer une bibliothèque

Réaliser des produits

Gérer des identifiants

Gérer les données de la recherche

CoopIST propose une série de fiches aux données.

Les données de la recherche s'insèrent dans les plans de recherche pour décrire le processus de valorisation.

CoopIST propose également [Rédiger et publier un data](#) pour communiquer et diffuser les données.

CoopIST propose également [Rédiger et publier un data](#) pour communiquer et diffuser les données.

### Voir aussi

[Rédiger et publier un data paper dans une revue scientifique](#)

[Citer un jeu de données scientifiques](#)

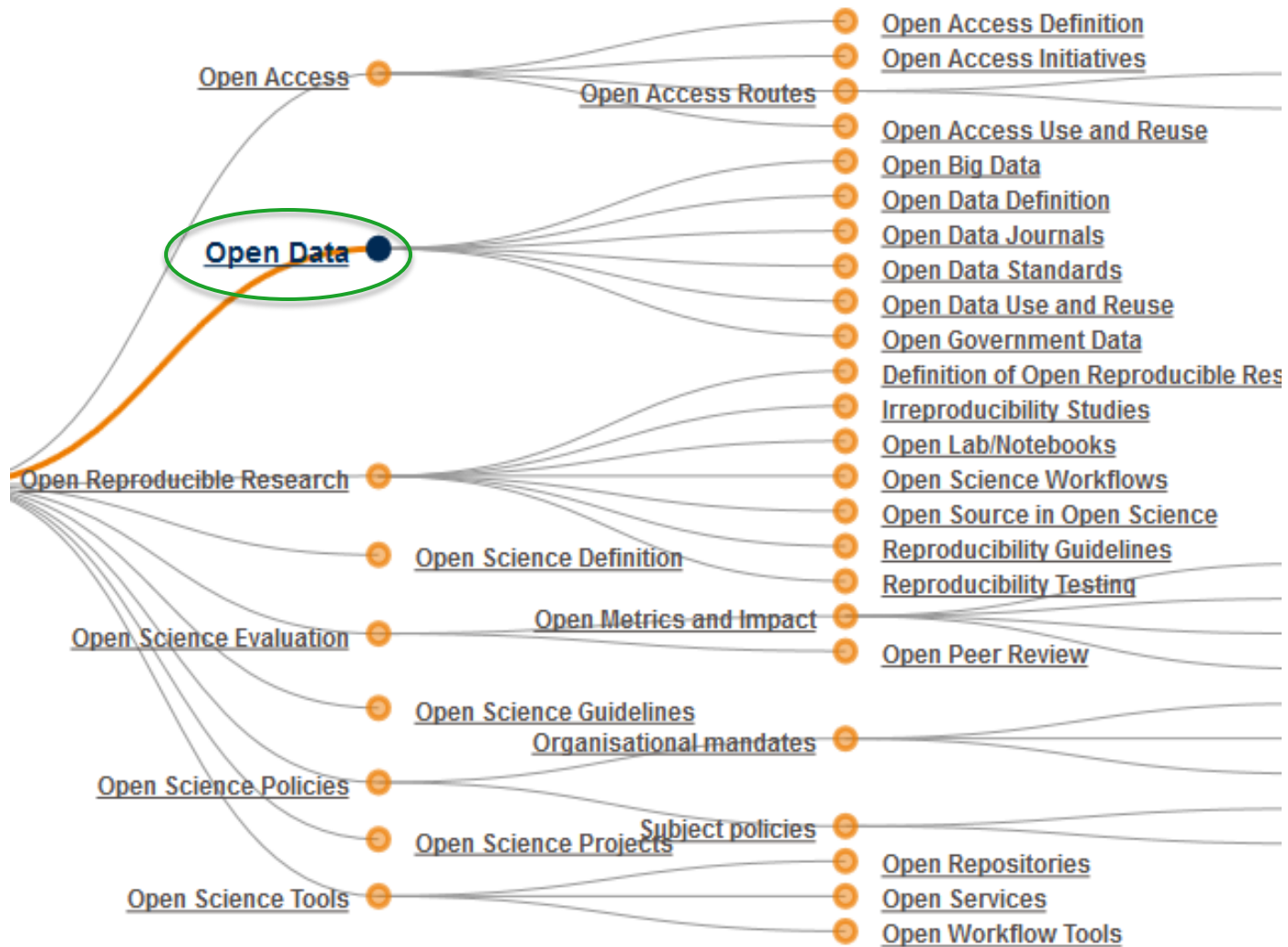
[Identifier et rechercher une publication ou un jeu de données par son DOI](#)

### Url courte

<http://coop-ist.cirad.fr/intro-gestion-donnees>

<https://coop-ist.cirad.fr/>

# Open data = une partie de l'Open science



<https://www.fosteropenscience.eu/foster-taxonomy/open-science>



# Open data

- **Des données de recherche ouvertes pour**
  - **Capitaliser** sur les résultats précédents (**qualité**)
  - Encourager la **collaboration**, éviter la dispersion des efforts (**efficacité**)
  - Accélérer l'**innovation** (**transfert**)
  - Impliquer les citoyens et la **société** (**transparence**)

Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

# Open data : les enjeux (1)

## Pour les chercheurs

### De nouvelles exigences et opportunités

- Le partage des données peut être **obligatoire** pour obtenir une **subvention** ou pour **publier** un article.
- Il permet aussi au chercheur de **valoriser ses données** et d'**accroître sa visibilité**.

### Exemple

*Les bénéficiaires d'un **financement H2020** doivent planifier le dépôt de leurs données...*

*Les données sont **citables** et la paternité du chercheur est reconnue.*

- **Plan de gestion de données** obligatoire pour Horizon 2020
- Accès aux données pour **publication** (obligatoire pour PLoS One)
- Valorisation des résultats (data papers), **citation**
- Promotion de l'auteur, peut amener des **collaborations**.



DORANum. Enjeux et bénéfices : fiche synthétique  
<http://doranum.fr/fiche-synthetique-enjeux-et-benefices/>

# Open data : les enjeux (2)



## Pour la science

### De nouvelles perspectives

Le partage des données permet aux chercheurs de s'appuyer sur un **socle plus important** et ouvre la voie à de **nouvelles méthodes d'investigations**.

*Utiliser des bases de données de référence, approcher de différentes façons un même jeu de données...*

### Une meilleure validation

- L'accès aux données facilite la **validation des articles** par les comités de lecture des revues
- Leur **réutilisation** peut amener un réexamen des résultats initiaux.



DORANum. Enjeux et bénéfices : fiche synthétique  
<http://doranum.fr/fiche-synthetique-enjeux-et-benefices/>



# Open data : les enjeux (3)



## Pour la société

La réutilisation des données permet **un meilleur usage de l'argent public** et **dynamise les relations entre la recherche et la société.**

Implication des citoyens dans le cadre des *sciences participatives*.

- Plus de **transparence** renforce la confiance des citoyens
- Exploitation de l'open data et du **big data** peut éclairer la société sur ses évolutions récentes.



DORANum. Enjeux et bénéfices : fiche synthétique  
<http://doranum.fr/fiche-synthetique-enjeux-et-benefices/>

# Données de la recherche : définition

- L'Organisation de coopération et de développement économiques (OCDE) définit les « **données de la recherche** » comme des *enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche.*
- L'ensemble des **données scientifiques produites**, d'origine numérique ou converties **sous forme numérique**, correspond à une **grande variété de supports** : textes, sons, images fixes et animées, multimédia, bases de données, logiciels, etc. On parle plus généralement de **ressource numérique scientifique**.

# Nature et granularité

## Données d'observation

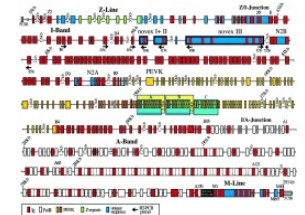
Données capturées en temps réel généralement uniques et impossibles à reproduire. Elles ont vocation à être conservées de façon pérenne.



enquêtes

## Données expérimentales

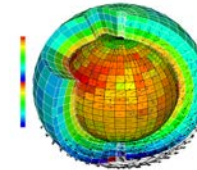
Données obtenues à partir d'équipements en laboratoire, suivant une méthodologie bien définie. Potentiellement reproductibles, mais à des coûts parfois prohibitifs.



Séquence de gènes

## Données de simulation

Données générées à partir de modèles.



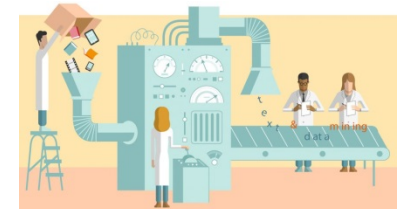
Modèle climatique

## Données dérivées ou compilées

Données résultant du traitement ou de la combinaison de données «brutes».

## Données de référence

Collection de jeux de données qui ont été revus, annotés et mis à disposition par les pairs.



Data mining – text mining

# Quelles données sont concernées ?

Toutes les données **brutes, traitées et analysées** sont susceptibles d'avoir un intérêt pour des utilisateurs.

<b>Données brutes</b>	Données <b>recueillies, non organisées, non mises en forme</b> . Elles peuvent intéresser un public susceptible de les <b>traiter</b> .
<b>Données traitées</b>	Données <b>corrigées, calculées, réorganisées...</b> Elles peuvent intéresser un public susceptible de les <b>analyser</b> .
<b>Données analysées</b>	Données telles qu'apparaissant dans les <b>publications</b> . Elles peuvent intéresser un public susceptible de les <b>comprendre</b> .

Doivent faire l'objet d'une attention particulière :

- les **données à caractère personnel** (nom, prénom, adresse de messagerie, adresse postale, ... dans le cadre *d'enquêtes, colloque, liste de contacts*)
- les **données relevant de la propriété intellectuelle et industrielle**
- les **données dites sensibles** (origines, opinions...)

# Qu'est-ce qu'une donnée sensible ?

Données sensibles > font apparaître, directement ou indirectement, les **origines raciales ou ethniques**, les **opinions politiques, philosophiques ou religieuses** ou l'**appartenance syndicale** des personnes

ou > sont relatives à leur **santé** ou à leur **vie sexuelle**.



Par principe leur collecte et leur traitement sont interdits.

Cependant, dans la mesure où la finalité du traitement l'exige, ne sont pas soumis à cette interdiction :

- Les traitements pour lesquels la personne concernée a donné son **consentement express**
- Les traitements justifiés par un **intérêt public après autorisation** de la CNIL ou décret en Conseil d'État

La collecte et le traitement de ces données doivent dans ces hypothèses être **justifiées au cas par cas au regard des objectifs poursuivis**.

# Qu'est-ce qu'un jeu de données ?

Un jeu de données (*dataset*) est l'agrégation d'enregistrements de données organisés pour former un **ensemble cohérent**. Les jeux de données numériques sont formatés de telle sorte qu'ils soient communicables, interprétables et adaptés à un traitement informatisé.



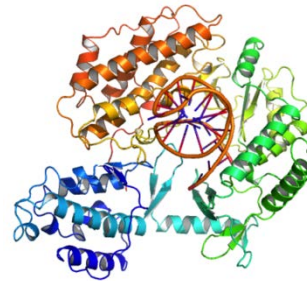
*Météo*



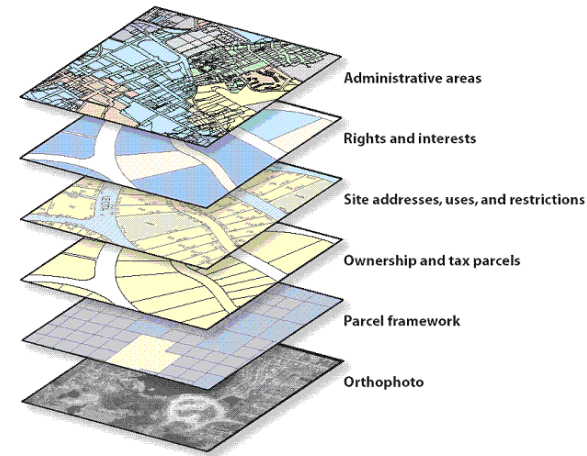
*Cartographie d'acteurs*



*Séries temporelles*



*Biologie cellulaire*



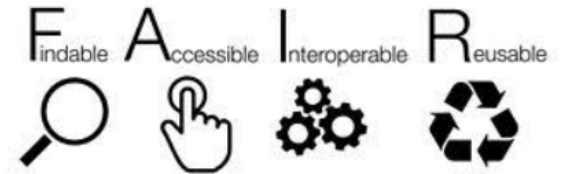
*Couches SIG*



# Gérer les données d'un projet

- **Choix du « grain » des jeux de données**
  - **Par discipline** = cohérence technique  
même type/nature/méthodes  
mêmes métadonnées
  - **Par sites, pays, workpackage** = cohérence scientifique
    - Dépend du projet
    - Métadonnées communes
- **Prendre en compte**
  - La **valeur scientifique** et le **potentiel de réutilisation**
  - Le **devenir envisagé** : publications / entrepôt

# Les principes « FAIR »



**Les données doivent être :**

- **Findable**

S'assurer que les données sont identifiables et localisables

✓ *Métadonnées*  
✓ *Identifiant*

- **Accessible**

Rendre ces données facilement accessibles par défaut

✓ *protocole standard*

- **Interoperable**

Veiller à ce que les données générées soient interoperables pour permettre le partage et la réutilisation

✓ *Standards*  
✓ *Vocabulaires*

- **Reusable**

Prendre les mesures nécessaires à la réutilisation la plus large possible des données.

✓ *Licence*

# Décrire ses données

- Avec des **métadonnées** (éléments descriptifs structurés)
  - Différents standards de métadonnées
    - Génériques (ex : Dublin Core)
    - Disciplinaires (ex : EML, DDI)
- Avec la **documentation associée**
  - Méthodes, protocoles, plan échantillonnage → fichier « Lisez moi »
  - Dictionnaires des variables, unités de mesure, abréviations
  - Equipement utilisé, méthode de calibration, contrôles
  - Questionnaires, guides d'enquêteurs, technique d'anonymisation
  - Schéma de la base de données, fichiers de syntaxe
  - Liens vers les articles publiés à partir de ces données



Richesse des métadonnées + documentation complète  
→ assure la compréhension et la réutilisation des données

# Décrire ses données (2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Case ID	1a	1b	1c	2	3	4a	4b	4c	4d	4e	4f	4g	4h	4i	4j	4k	4l	4m	4n	5a
92	044	2	10		5	1	0	0	0	0	0	1	0	0	0	0	0	0	0		5
93	043	2	10		3	1	0	0	0	0	1	0	0	0	0	0	0	0	1	I am a c	4
94	042	2	10		4	1	0	0	0	0	1	1	0	0	0	0	0	0	1	Tradition	5
95	041	2	10		2	1	0	0	0	0	1	0	0	0	0	0	0	0	0		5
96	040	2	10		2	1	1	0	0	0	0	1	0	0	1	0	1	0	0		5
97	039	2	10		3	2	0	0	0	0	0	1	0	0	0	0	0	0	0		5
98	038	2	4		5	2	0	0	0	0	0	1	0	0	0	0	0	0	0	Keeping	5
99	037	2	2		2	2	0	1	0	0	0	1	0	0	0	0	0	0	0	Keeping	5
100	036	2	4		2	1	0	1	0	0	0	0	0	0	0	0	0	0	0		5
101	035	2	4		2	1	0	0	0	0	0	1	0	0	0	0	0	0	0		5
102	034	2	2		2	1	0	0	0	0	1	1	0	0	0	0	0	0	0		5
103	033	2	2		1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	I work in	5
104	032	2	4		3	2	1	0	0	0	1	1	0	0	1	0	1	0	0	Educate	5
105	031	2	4		4	2	1	0	0	0	0	0	1	0	0	0	0	0	0	Keeping	2
106	030	2	1		4	2	1	0	0	1	1	0	0	0	1	1	1	0	0	Keeping	5
107	029	2	2		3	1	0	0	0	0	0	1	0	0	0	0	0	0	0	Selling	5
108	028	2	4		2	1	1	0	0	0	0	1	0	0	0	0	1	0	0	I am a p	5
109	027	2	1		2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	Keeping	5
110	026	2	2		5	2	0	0	0	0	1	1	0	0	0	1	0	0	0	Counse	5
111	025	2	1		4	2	0	0	0	0	0	1	0	0	0	0	0	0	0	Military	5
112	024	2	11		3	2	0	0	0	0	1	0	0	0	0	0	0	0	0		5
113	023	2	11		4	1	1	0	0	0	0	1	0	0	0	0	1	0	0		5
114	022	2	11		4	2	1	0	0	0	1	0	0	0	0	0	0	0	0	Keeping	5
115	021	2	11		2	2	0	0	0	0	1	1	0	0	0	0	0	0	1		5
116	020	2	11		3	2	0	0	0	0	1	0	0	0	0	0	0	0	0	camera	4
117	019	2	11		5	2	0	0	0	0	1	1	0	0	0	0	0	0	0		5
118	018	2	11		3	1	0	0	0	0	1	0	0	0	0	0	0	0	0	I purcha	5
119	017	2	11		4	2	0	0	0	0	1	1	0	0	0	1	0	0	0		5



Est-ce suffisant pour comprendre et réutiliser les données ???

# Les métadonnées (1)

Les métadonnées sont des « *données qui décrivent des données* » :

- **Information** structurée associée à un "objet", un document ou un jeu de données
- **Documentation** qui permet à l'utilisateur de comprendre, de comparer et d'échanger le contenu du jeu de données décrit



Il est conseillé de produire les métadonnées au **moment de la collecte ou de la création** des données plutôt qu'à posteriori. Les métadonnées seront **complétées tout au long du cycle de vie** des données.

Des objets discernés par tous



*Un objet sans étiquette n'est connu que de son auteur*

Métadonnées



# Les métadonnées (2)

## Métadonnées

Eléments  
Structurels

Valeurs  
descriptives  
administratives

## Exemples de métadonnées

Elt structurel

titre :  
résumé :

auteur :  
format :  
accès :

Valeurs descriptives

Inventaire de dispositifs  
Dispositifs forestiers de type  
« parcelles permanentes »  
suivies depuis 1982

Valeurs administratives

S. Gourlet  
PostgreSQL  
Contrôlé sur demande

## Stockage numérique

```
<?xml version="1,0" encoding="utf-8?>
<metadata>
  <dc:title>Inventaire de dispositifs</dc:title>
  <dc:description>
    <dct:abstract>
      Dispositifs forestiers de type « parcelles
      permanentes » suivies depuis 1982
    </dct:abstract>
  </dc:description>
  <dc:creator>S. Gourlet</dc:creator>
  <dc:format>PostgreSQL</dc:format>
  <dc:rights>
    <dct:accessRights>Contrôlé sur
    demande</dct:accessRights>
  </dc:rights>
</metadata>
```



# Des métadonnées pour

Faciliter la découverte des données	+ Identification de la ressource (DOI)
	+ Description de la ressource
Faciliter la gestion et l'archivage	
Faciliter la gestion des droits	+ Droits de propriété intellectuelle
	+ Droits d'accès et d'utilisation
Faciliter la réutilisation des données	+ Informent sur le contenu de la ressource
	+ Informent sur la structure du jeu de données
	+ Informent sur la qualité
Faciliter l'interopérabilité	+ Format d'échange commun (XML)
	+ Vocabulaires contrôlés (thésaurus, ontologies...)
	+ Syntaxe définie dans des schémas génériques

# Les standards de MD (1)

- **Pourquoi utiliser un standard ?**
  - Facilite la recherche
  - Constitue un **modèle** pour
    - Ne pas oublier des informations importantes
    - Remplir correctement chaque champ
- **L'enjeu** = quelle que soit la machine...
  - **Retrouver** vos ressources
  - Les **protéger** (propriété intellectuelle)
  - Les **diffuser** largement
- **Adoptez un standard dès le début du projet**

# Les standards de MD (2)

- **Standard**

= ensemble de **recommandations** développées et préconisées par un groupe représentatif d'utilisateurs.

- **Norme**

= document établi par un consensus et **approuvé par un organisme reconnu**, qui fournit, pour des usages communs et repérés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné.

# Critères de choix d'un standard

- **Le type de ressources**

Champs différents pour une publication, une vidéo, une image...

- **Le domaine scientifique**

Pour représenter finement des données en fonction des pratiques de votre communauté

- **L'entrepôt choisi pour le dépôt**

Chaque entrepôt propose des standards

- **Les exigences de l'éditeur**

Si vous envisagez de publier un data paper

# Le choix d'un entrepôt

- A faire **au plus tôt**
  - Adopter dès le début du projet un ou des standards de MD pour enregistrer les données
- Suivre les **recommandations**

Un entrepôt est parfois imposé par l'institution, le financeur du projet, l'éditeur de l'article...
- Consulter le **répertoire re3data.org**



# MD génériques : le Dublin Core

## 15 métadonnées de base

DC Element	Notes
<b>Title</b>	Title of Data Collection
<b>Creator</b>	Authoring Entity of Data Collection
<b>Subject</b>	Keyword(s)
<b>Description</b>	Abstract
<b>Publisher</b>	Producer of Data Collection
<b>Contributor</b>	
<b>Date</b>	Production Date - Data Collection
<b>Type</b>	Kind of Data
<b>Format</b>	Type of File
<b>Identifier</b>	ID Number - Data Collection
<b>Source</b>	Sources - Used for Data Collection
<b>Language</b>	
<b>Relation</b>	Other Study Description Materials
<b>Coverage</b>	Time Period Covered
	Country
	Geographic Coverage
<b>Rights</b>	Copyright - Data Collection

Norme ISO 15836

Minimum d'informations structurées décrivant les données

Lecture par l'homme et par les machines

**Standard générique** s'appliquant à tous types de ressources et largement utilisé

Inclut des *recommandations* pour le format de certains contenus  
Ex. pour les dates : Norme ISO-8601 (AAAA-MM-JJ)



# Standards de MD disciplinaires

Quelques exemples :

## ■ Géolocalisation

- Geospatial metadata standard, norme ISO 19115
- Directive européenne INSPIRE (2007)

## ■ Ecologie, Biodiversité

- Ecological Metadata Language (EML)
- Darwin Core

## ■ Sciences Humaines et Sociales

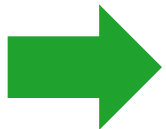
- Data Documentation Initiative (DDI)

Des répertoires pour trouver un standard adapté à vos besoins :

DCC Digital Curation Centre : <http://www.dcc.ac.uk/resources/metadata-standards>

FAIRsharing : <https://fairsharing.org/standards/>

RDA Metadata Directory : <http://rd-alliance.github.io/metadata-directory/standards/>



# Standards de MD disciplinaires

## Physical Science

Space science Remote Sensing Geography Astrophysics Nuclear and Particle  
Physics Solar physics Multi-disciplinary Astronomy Materials Science  
Meteorology Geology Physics Biochemistry Crystallography Molecular biology  
Bioinformatics Chemistry Geoscience

## Metadata Standards

### AVM - Astronomy Visualization Metadata

A standard defining discovery metadata for fully rendered astronomical imagery.

### CIF - Crystallographic Information Framework

An extensible standard file format and set of protocols for the exchange of crystallographic and related structured data.

### CSMD-CCLRC Core Scientific Metadata Model

A study-data oriented model that captures high-level information about scientific studies and the data that they produce, primarily tailored for the physical sciences.

### FITS - Flexible Image Transport System

Used by the astronomy community to originally describe telescope images, but is now a family of standards to describe multi-dimensional data including spatial, spectral and temporal dimensions and the scaling and distortions that may be present.

### International Virtual Observatory Alliance Technical Specifications

A set of specifications, including metadata standards, that enable the integration of many astronomical archives into an international virtual observatory.

### NeXus

NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data.

## Observations and Measurements

# Pas de standard dans mon domaine ?

Quelques pistes si vous ne trouvez pas de standard approprié...

- Cherchez un **entrepôt spécialisé** et vérifiez les MD proposées
- Cherchez si des **articles** ont été publiés sur des standards en cours de définition par la communauté
- Explorez les **sites des sociétés savantes**, groupes thématiques, réseaux, projets internationaux...
- Cherchez des **data papers** dans la discipline, vérifiez les entrepôts recommandés par les **éditeurs** de revues spécialisées, voyez les standards utilisés

# Bien décrire son jeu de données

Titre  
significatif

Description

Fichier  
readme  
pour  
expliquer  
les  
données  
(fichiers,  
variables,  
etc.)

April 15, 2018

Dataset Open Access

## Figures in Open Access Scientific Publications - Underlying Data

10 Sohmen, Lucia; 10 Charbonnier, Jean; 10 Blümel, Ina; 10 Wartena, Christian; 10 Heller, Lambert

This publication contains data for a statistical analysis of an OA article corpus. The underlying dataset consists of over 1 million open access articles from different publishers (Copernicus: 9592; Springer:78418; Hindawi: 147848; Frontiers: 57621; PMC (aggregator): 747839)

Preview

Files (44.4 MB)

Name	Size	
<a href="#">AffiliationNetwork-Edges.csv</a>	26.4 kB	<a href="#">Preview</a> <a href="#">Download</a>
md5:c7f9fe1c03a5ce5625cb286a74d0c774		
<a href="#">AffiliationNetwork-Nodes.csv</a>	1.6 kB	<a href="#">Preview</a> <a href="#">Download</a>
md5:acba18fcff5c42957e8f53b3ce49072e		
<a href="#">affiliations_visualized_network.jpg</a>	832.8 kB	<a href="#">Preview</a> <a href="#">Download</a>
md5:42f56912f18ed37a00aaf43dd3899ed6		
<a href="#">CaptionLength.csv</a>	1.3 kB	<a href="#">Preview</a> <a href="#">Download</a>
md5:17f3eed41e1365cb92835e62b9012dbf		
<a href="#">ChartsVSIImages.csv</a>	2.7 kB	<a href="#">Preview</a> <a href="#">Download</a>
md5:01b83b4cda7112e1b395b7c385a307a5		
<a href="#">CitationCorrelation.csv</a>	43.6 MB	<a href="#">Preview</a> <a href="#">Download</a>
md5:85af1b1c96b63f9441b0a4f45039bb94		
<a href="#">readme.txt</a>	1.2 kB	<a href="#">Download</a>
md5:9b924147312ced94d2452466e2a1a32c		

Attention aux  
noms des fichiers

Indexed in

OpenAIRE

Publication date:

April 15, 2018

DOI:

DOI 10.5281/zenodo.1218724

Keyword(s):

Open Access Scientific Figures Statistical Analysis

License (for files):

[Creative Commons Attribution 4.0](#)

Versions

Version 1 10.5281/zenodo.1218724 Apr 15, 2018

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.1218723. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share



Cite as

Sohmen, Lucia, Charbonnier, Jean, Blümel, Ina, Wartena, Christian, & Heller, Lambert. (2018). Figures in Open Access Scientific Publications - Underlying Data [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1218724>

Start typing a citation style

Quelques  
mots-clés

On  
obtient  
une  
citation  
avec DOI

## TP 1 – Évaluez la description d'un jeu de données

- Exemples de jeux de données déposés dans **Zenodo**  
<https://zenodo.org/>
- Chaque participant traite un jeu de données dans la liste fournie (sélections thématiques)

### EXERCICE

- **Analysez leur description** (titre, résumé, mots-clés...) en tant que potentiel utilisateur : sont-ils faciles à retrouver, à comprendre, à réutiliser ?...

# TP 1 : exemples de jeux de données (1)

## HISTOIRE

<https://zenodo.org/record/1159078>

<https://zenodo.org/record/835588>

<https://zenodo.org/record/1056832>

<https://zenodo.org/record/1183367>

## SCIENCES SOCIALES ET ECONOMIQUES

<https://zenodo.org/record/48799>

<https://zenodo.org/record/1174366>

<https://zenodo.org/record/375612>

<https://zenodo.org/record/56160>



# TP 1 : exemples de jeux de données (2)

## SCIENCES DE LA TERRE ET DE L'UNIVERS

<https://zenodo.org/record/1220566>

<https://zenodo.org/record/12873>

<https://zenodo.org/record/1145822>

<https://zenodo.org/record/7515>

## SCIENCES DE LA VIE

<https://zenodo.org/record/16837>

<https://zenodo.org/record/1237738>

<https://zenodo.org/record/1230390>

<https://zenodo.org/record/12347>

## TP 2 – Déposez un jeu de données

- Déposez un des jeux de données fournis sur le site **Dataverse "demo site"** : <https://demo.dataverse.org/>
  - Se créer un compte : Sign Up (en haut à droite)
  - Pour déposer : Add Data
  - Déposer le fichier, saisir les MD obligatoires, sauvegarder
  - Compléter le dépôt : onglet Metadata / Add and Edit Metadata
- MD de citation et géospatiales : norme **Dublin Core**
- MD disciplinaires : **3 standards** proposés
  - Sciences humaines et sociales > DDI
  - Sciences de la vie > ISA-Tab
  - Astronomie > Virtual Observatory Metadata

# Le dépôt dans Dataverse (1)

Champs \*  
obligatoires

\*Asterisks indicate required fields

Citation Metadata ^

**Title \***

**Author \*** **Name \***  **Affiliation**

**Identifier Scheme**  **Identifier**

**Contact \*** **Name**  **Affiliation**

**E-mail \***

**Description \***

**Subject \***

☐ Agricultural Sciences  
☐ Medicine, Health and Life Sciences  
☐ Arts and Humanities  
☐ Astronomy and Astrophysics  
☐ Other  
☐ Computer and Information Science

**Keyword**

**Term**

**Vocabulary**

Champs  
répétables

13 grands  
thèmes

Possibilité  
d'utiliser un  
vocabulaire  
contrôlé (à  
spécifier)

# Le dépôt dans Dataverse (2)

Demo Dataverse > Deuxième test Bordeaux de dépôt dans le site démo Dataverse


 Metrics

0 Downloads

 Contact


 Share

 Publish

 Edit ▾



## Deuxième test Bordeaux de dépôt dans le site démo Dataverse Draft Unpublished

Barale, Martine, 2018, "Deuxième test Bordeaux de dépôt dans le site démo Dataverse", <https://doi.org/10.5072/FK2/UDPGDD>, Demo Dataverse, DRAFT VERSION, UNF:6:bsVd14HsOndhVkOmQKR5MA== 

 Cite Dataset ▾

 Learn about Data Citation Standards.

### Description

fjqmsdlkja aei mjqeoit jiq ezit ròuò qoeo

### Subject

Arts and Humanities; Astronomy and Astrophysics


**Files**


Metadata

Terms

Versions

1 File

 Upload Files


 Edit Files ▾




**boites\_archives.tab**

Tabular Data - 137 bytes - May 17, 2018 - 0 Downloads

3 Variables, 13 Observations - UNF:6:bsVd14HsOndhVkOmQKR5MA==

 Explore ▾

 Download ▾

# Le dépôt dans Dataverse (3)

Après avoir validé la description (Dublin Core), on peut renseigner des MD disciplinaires selon plusieurs standards proposés par Dataverse

Geospatial Metadata ▼

Social Science and Humanities Metadata ▼

Astronomy and Astrophysics Metadata ▼

Life Sciences Metadata ▼

Journal Metadata ▼

Save Changes Cancel

# Le dépôt dans Dataverse (4)

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Add + Edit Metadata

Citation Metadata ^

Dataset Persistent ID	doi:10.5072/FK2/UDPGDD
Title	Deuxième test Bordeaux de dépôt dans le site démo Dataverse
Author	Barale, Martine (Université de Bordeaux) - ORCID:
Contact	Use email button above to contact. Barale, Martine
Description	fjqmsdlkja aei mjqeoit jiq ezit rùoù qoeo
Subject	Arts and Humanities; Astronomy and Astrophysics
Related Publication	doi:
Depositor	Barale, Martine
Deposit Date	2018-05-17

Geospatial Metadata ^

Geographic Coverage	Angola
---------------------	--------

Life Sciences Metadata ^

Factor Type	Treatment Type
Organism	Oryza sativa
Measurement Type	cell counting

# Le dépôt dans Dataverse (5)

## Onglet Terms = licence d'utilisation

- Par défaut : licence Creative Commons "Public Domain" = CC0
- Sinon : choisir une autre licence (<https://creativecommons.org/choose/>)
- Copier/coller le code fourni → génère le logo de la licence à l'affichage

Save Changes Cancel

Terms

Terms of Use ^

**Waiver**

Datasets will default to a [CC0 public domain dedication](#). CC0 facilitates reuse and extensibility of research data. Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. If you are unable to give datasets a CC0 waiver you may enter custom Terms of Use for datasets.


☐ Yes, apply CC0 - "Public Domain Dedication" ☒ No, do not apply CC0 - "Public Domain Dedication"

**Terms of Use**

If you are unable to use CC0 for datasets you are able to set custom terms of use. Here is an example of a [Data Usage Agreement](#) for datasets that have de-identified human subject data.

```
<a rel="license" href="http://creativecommons.org/licenses/by/4.0/">
```

Terms of Use



## TP 3 - Recherchez des entrepôts et des standards de MD dans votre discipline

- Recherche d'entrepôts
  - <https://www.re3data.org/>
- Recherche de standards
  - <http://www.dcc.ac.uk/>
  - <https://fairsharing.org/>
  - <http://rd-alliance.github.io/metadata-directory/standards/>



# Quelques sites utiles

- H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, V3 (2016)  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- The FAIR Guiding Principles for scientific data management and stewardship  
Wilkinson, MD et al. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>
- Guiding principles for FAIR data publishing (V B1.0)  
<https://www.force11.org/fairprinciples>
- Guide to Social Science Data Preparation and Archiving  
<https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html>
- Guide to writing "readme" style metadata, Cornell University  
<https://data.research.cornell.edu/content/readme>
- DORANum (Cnrs)  
<http://doranum.fr/>
- Research data management e-learning platform. Gestion des données de recherche  
<http://www.researchdatamanagement.ch/fr/>
- REgistry of REsearch data REpositories  
<https://www.re3data.org/>
- Eléments de métadonnées du Dublin Core, Version 1.1: Description de Référence  
(traduction française de Anne-Marie Vercoustre, Inria)  
<https://who.rocq.inria.fr/Anne-Marie.Vercoustre/METADATA/DC-fr.1.1.html>

# Merci pour votre participation

Martine BARALE  
martine.barale@cirad.fr

## Remerciements

Cet atelier a été préparé avec l'aide précieuse de mes collègues

Sophie Fortuno

UMR Tetis, chef de projet "Gestion et valorisation des données scientifiques au Cirad"

Laurence Dedieu

DIST, éditrice en appui aux unités de recherche pour la gestion des données

Fabrice Bénédet

UMR Forêts et Sociétés, gestionnaire de données